An introduction to time series and time series models

A lecture from Applied Time Series Analysis for Ecologists

Link to online course at http://faculty.washington.edu/eeholmes/

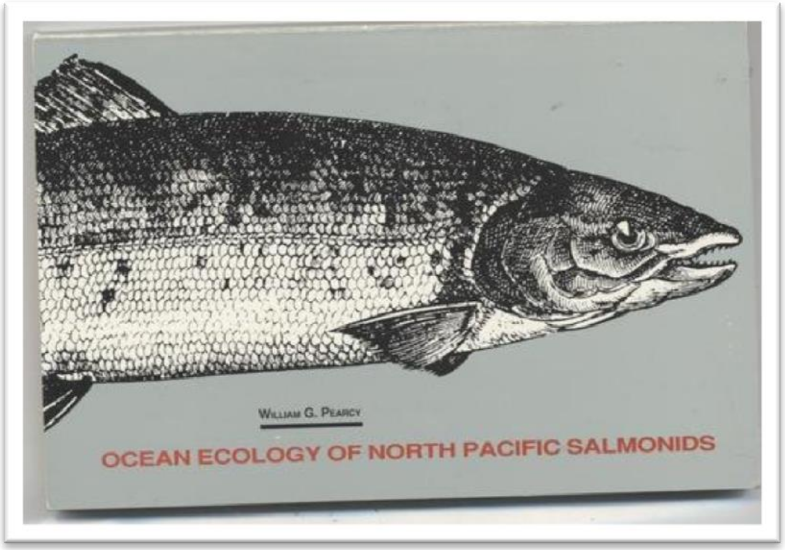
Eli Holmes, NOAA

Hyderabad, India

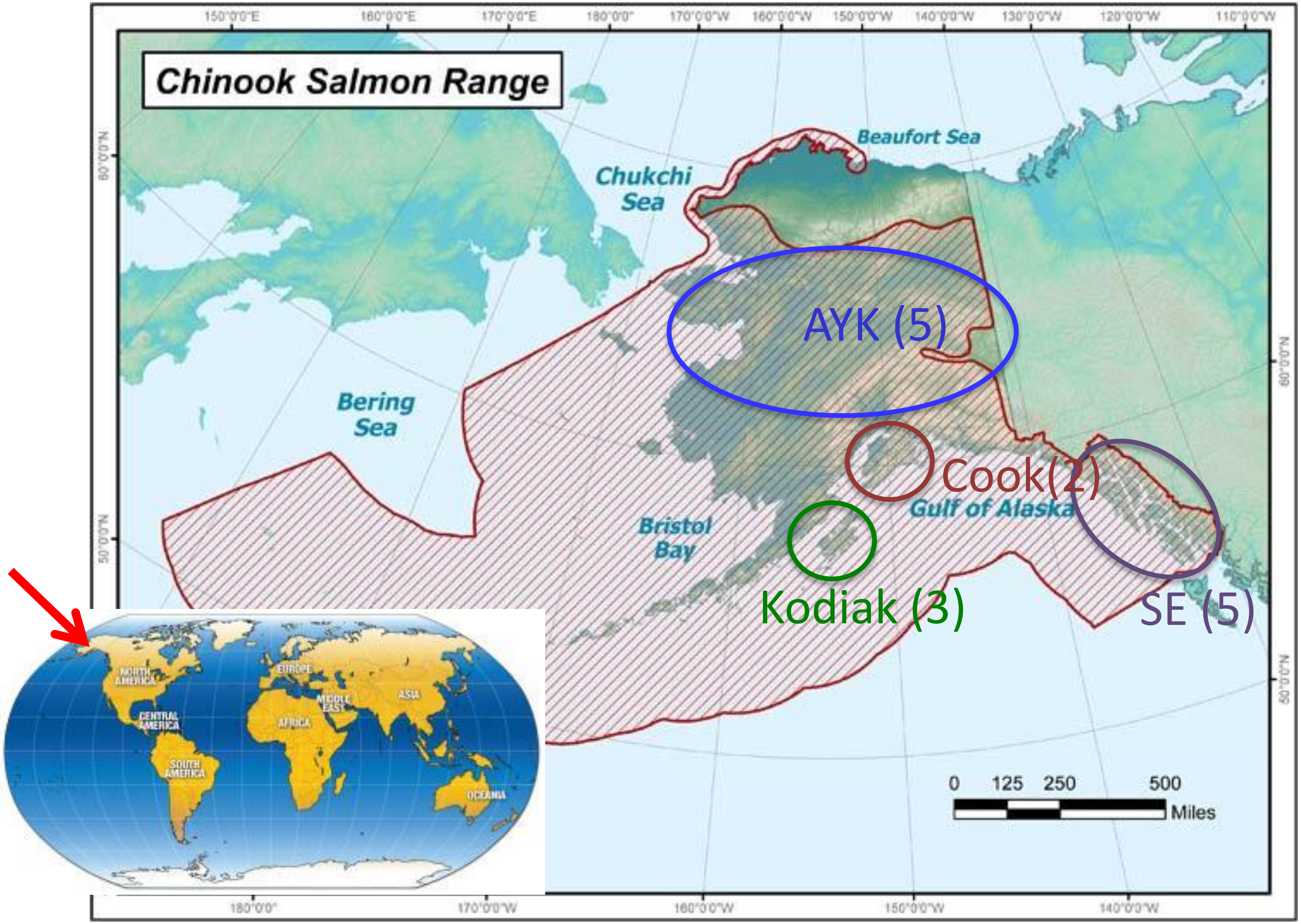
16-22 Sept 2015

An example of a study using time-series analysis work by Mark Scheuerell at NWFSC, Seattle

What evidence exists to support the hypothesis that large-scale ocean-climate drives fluctuations in Alaska salmon survival?

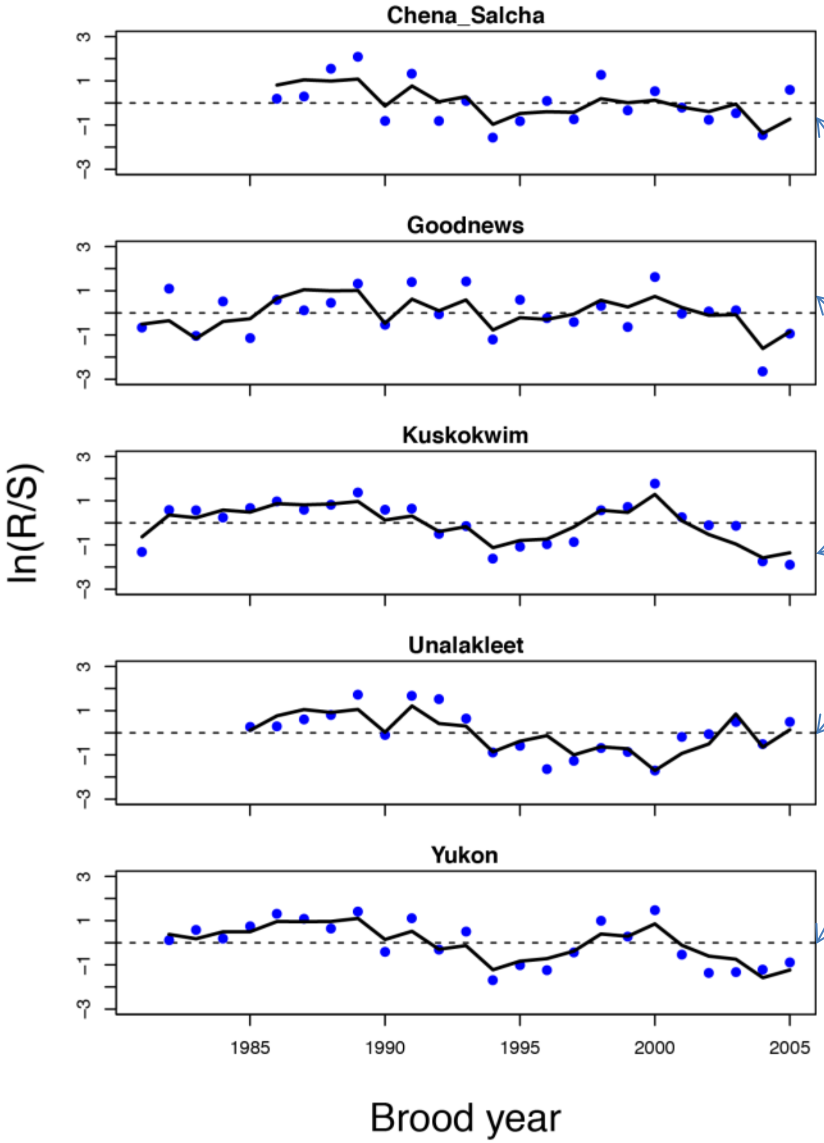


Alaska Chinook salmon



The data: time series of log recruits per spawner versus brood year for AYK region

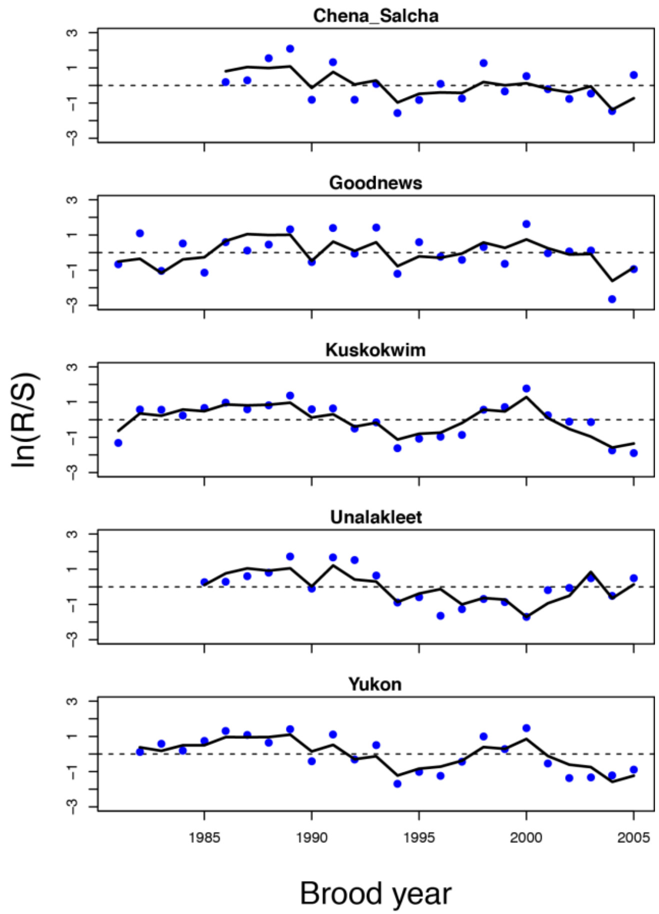
Estimate of survival



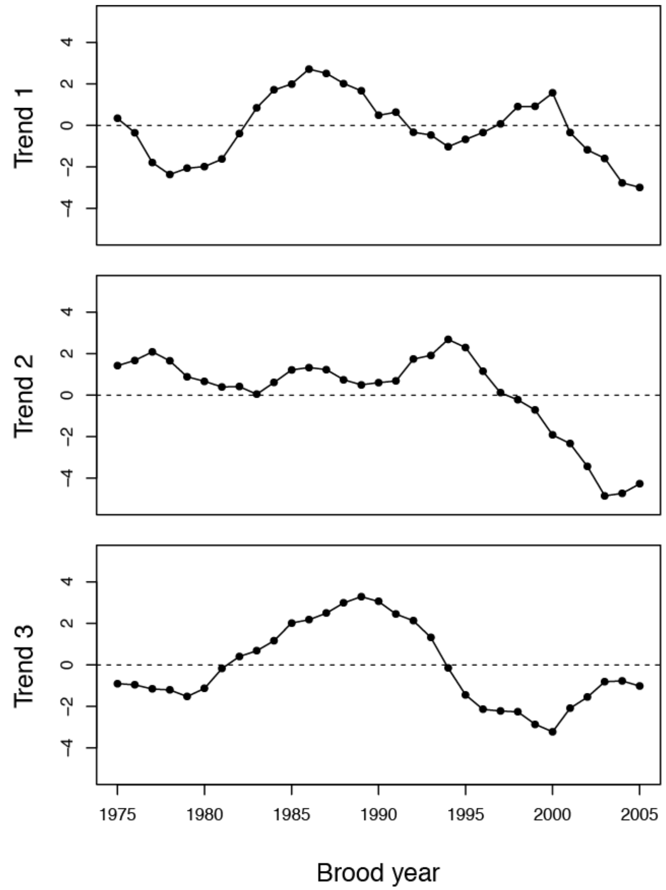
Different stocks/populations within one of the Alaska regions

The analysis
(Dynamic Factor Analysis)

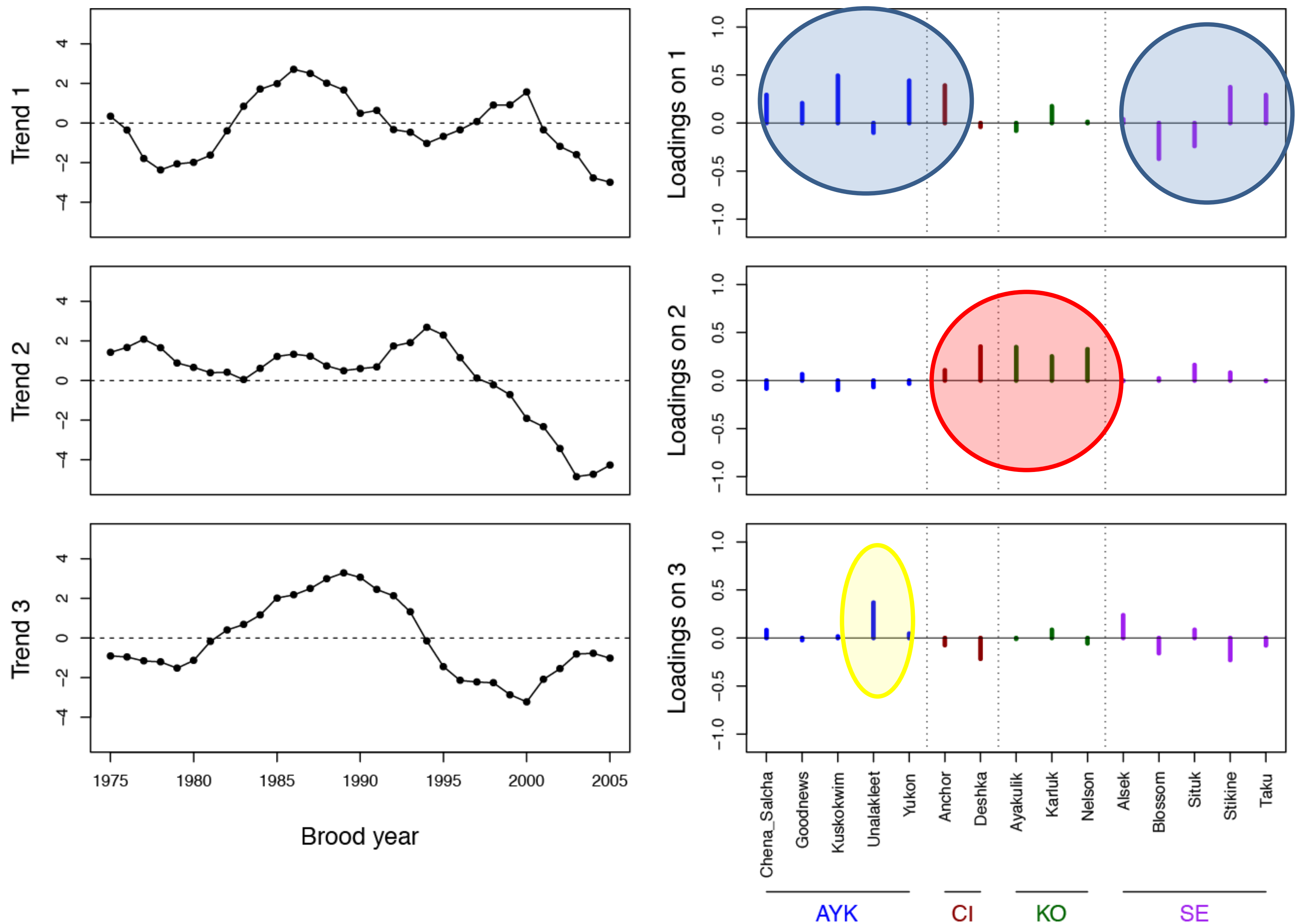
Raw Data: 15 time series
(5 of 15 shown)



Can be described by 3
overall patterns

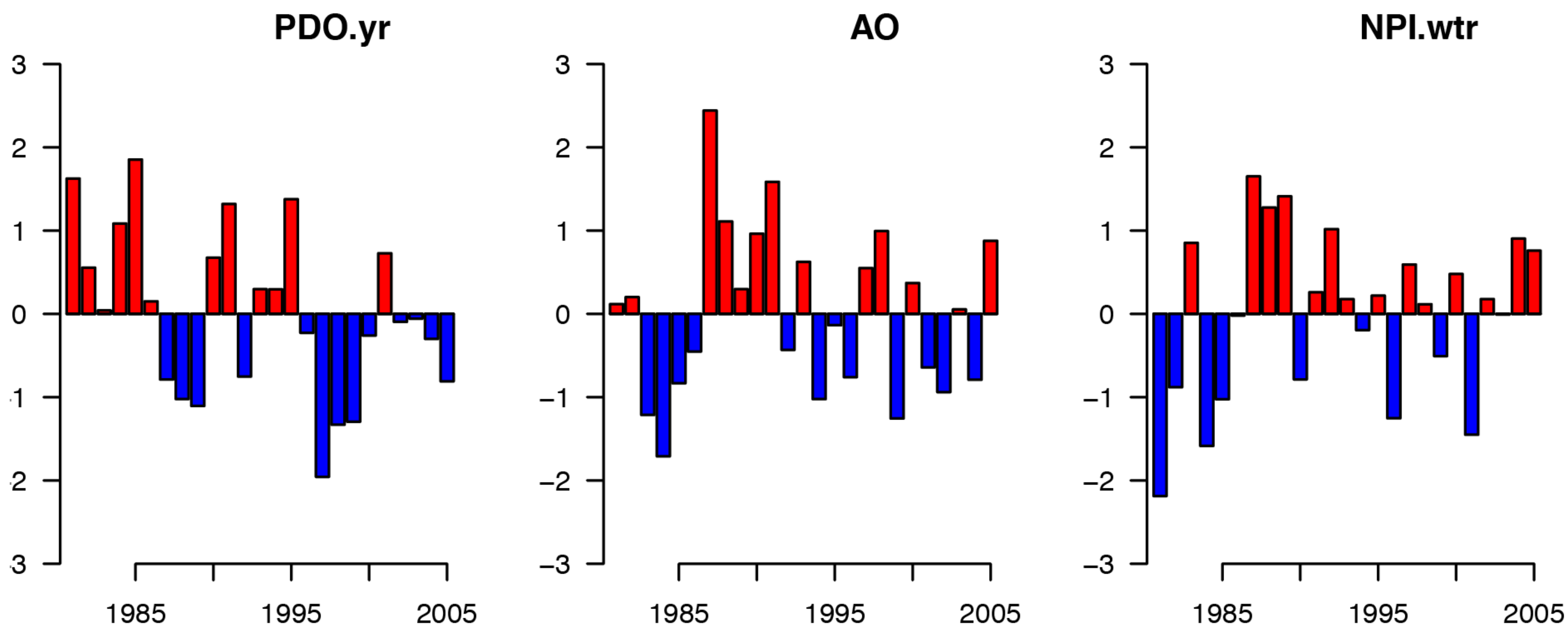


Results: 3 large-scale trends/drivers
The different stocks weight on these differently.



Next analysis looks at correlation between the overall trends and large-scale environmental indicators

- Pacific Decadal Oscillation
- Arctic Oscillation Index
- Aleutian Low Pressure
- North Pacific Index



Introduction to time-series analysis in R

- Characteristics of time series (ts)
  - What is a ts?
  - Classifying ts
  - Trends
  - Seasonality (periodicity)
  - Stationarity
- Time-series models
  - White noise
  - Random walks
  - Autoregressive (AR) models
  - Moving average (MA) models
  - ARMA models
- Diagnostics for time series
  - Autocorrelation functions (ACF)
  - Correlograms

What is a time series?

- A *time series* (ts) is a set of observations taken sequentially in time

- A ts can be represented as a set

  $\{x_t : t = 1,2,3,\ldots,n\} = \{x_1,x_2,x_3,\ldots,x_n\}$

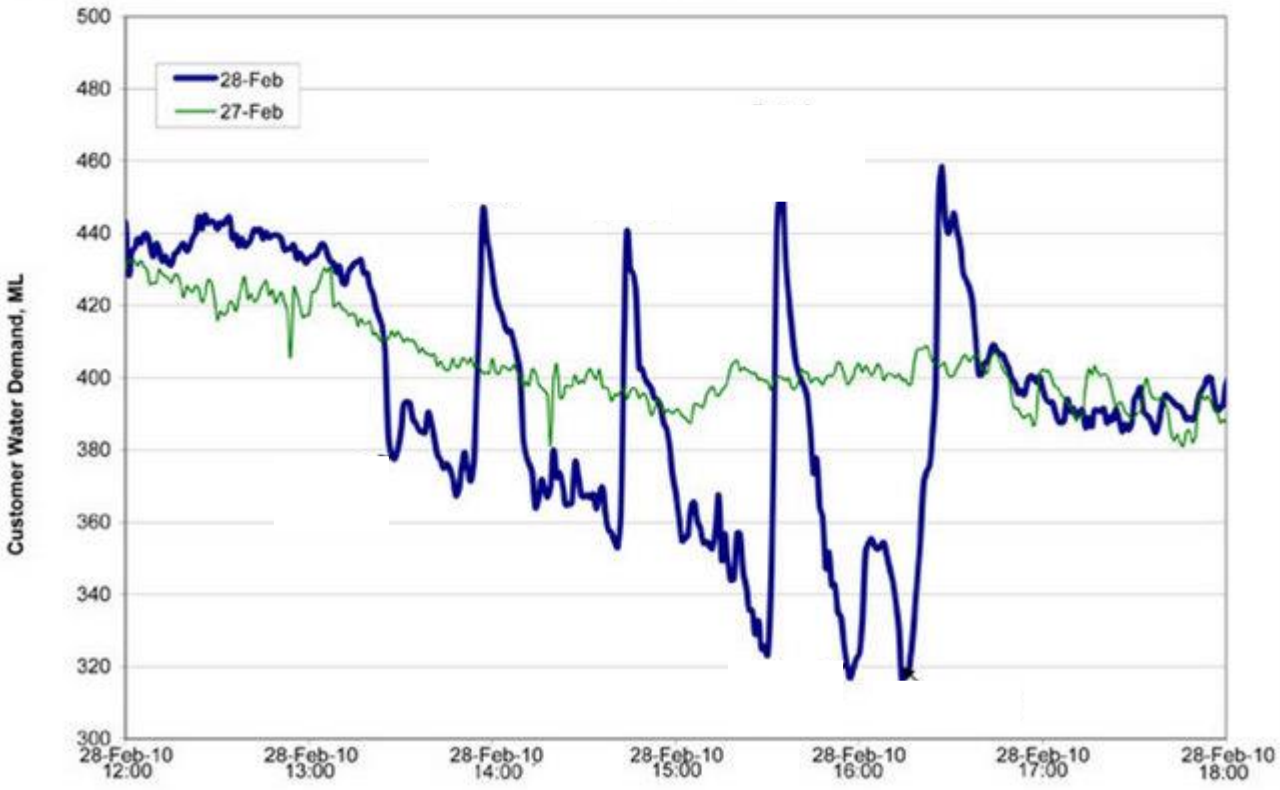
- For example,

  {10,31,27,NA,53,15}

- Univariate (e.g. total # of fish caught) or multivariate (e.g. # of each species caught)

Example of a time series of water usage typical versus during hockey championship game

Water Usage



Time of Day

How do we describe a time series?

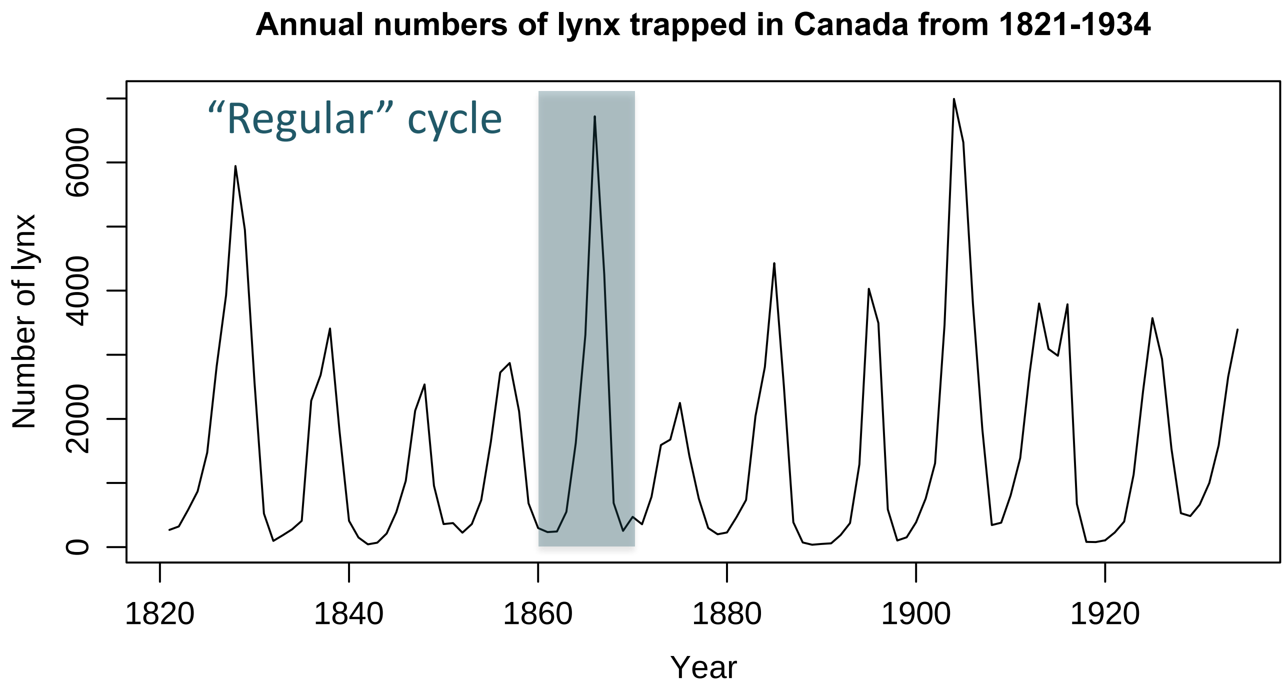
$$observation_t = \text{trend} + \text{cycle} + e_t$$

$$e_t = \text{a time series also}$$

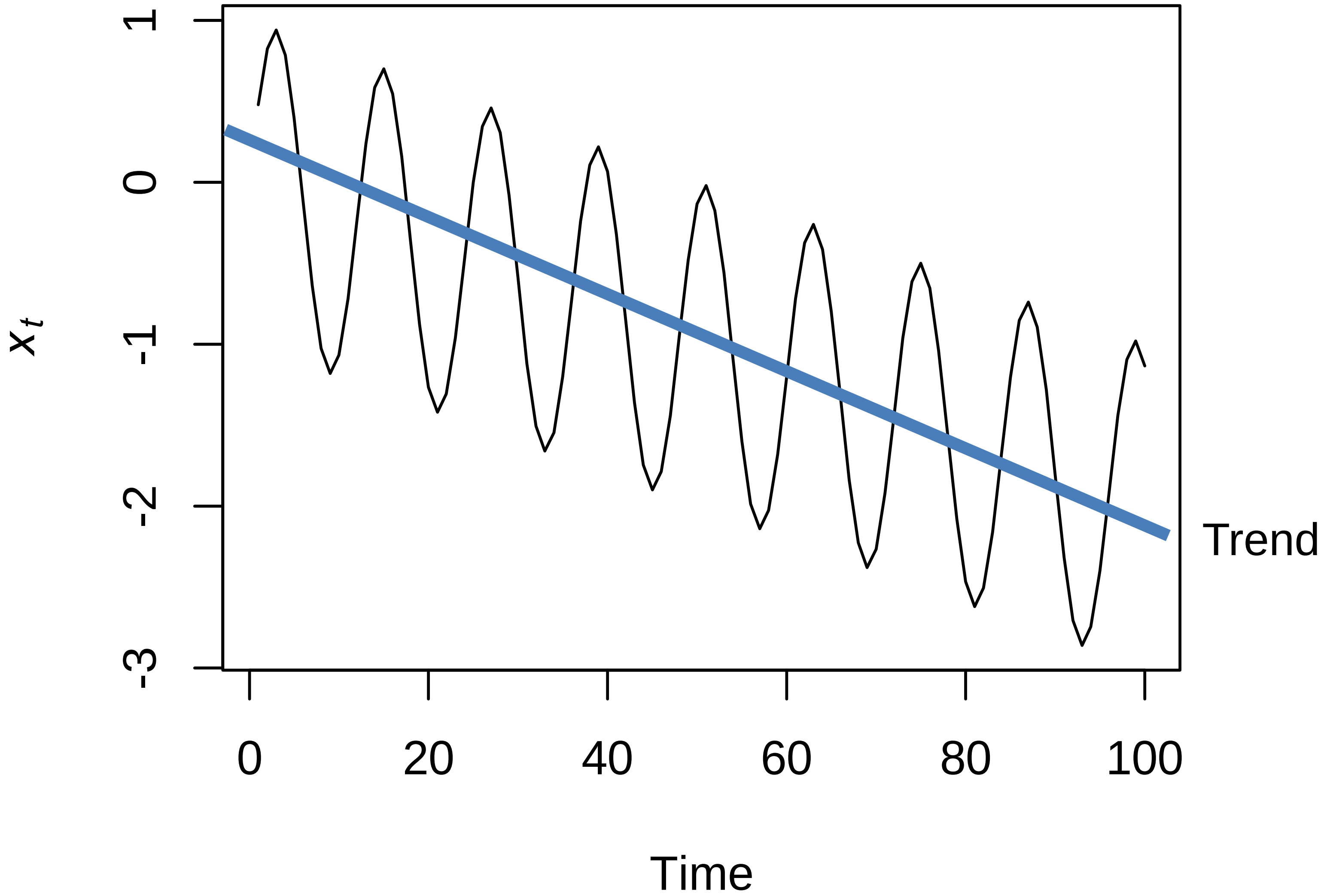
Often the objective is to estimate or describe the trend and cycle in a time series, but to do this we need to describe/model the e_t.

Other times the objective is to model the e_t since we are trying to understand what drives the year-to-year (month-to-month) variation.

Cycles or seasonality in a time series



Trend in a time series



Stationarity of time series

- *Stationarity* describes a particular statistical properties of a time series.

- In general, a time series is said to be stationary if there is
  1) no systematic change in mean or variance,
  2) no systematic trend up or down, and
  3) no periodic variations or seasonality

We typically remove the trend and cycles and treat e_t as stationary.

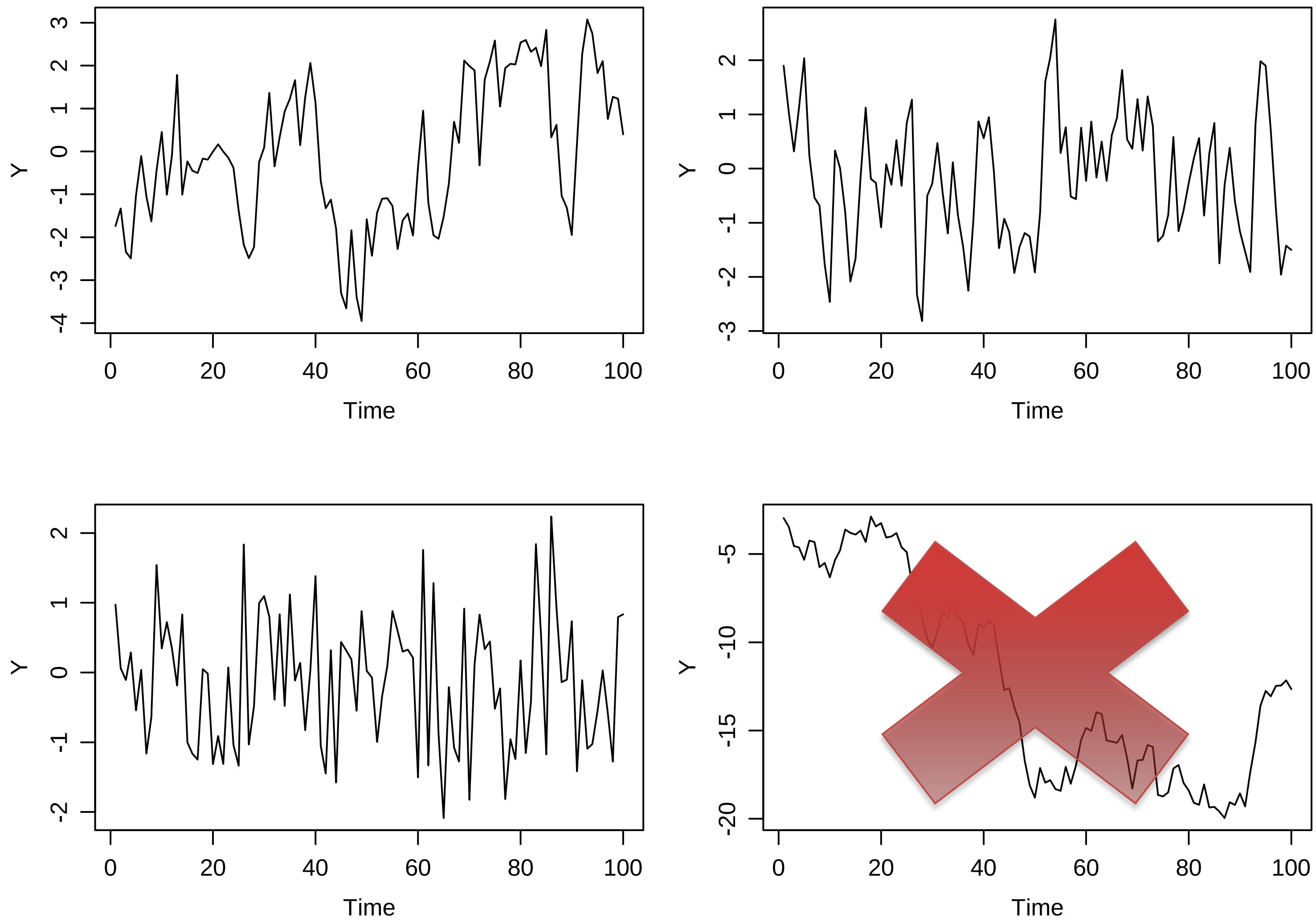
Describing a time series: classical decomposition

- *Classical decomposition* of an observed time series is a fundamental approach in time series analysis
- The idea is to decompose a time series $\{x_t\}$ into a trend, a seasonal component, and a remainder (e_t)

$$observation_t = \text{trend} + \text{cycle} + e_t$$

$$e_t = \text{a time series also}$$

Which of these are stationary?



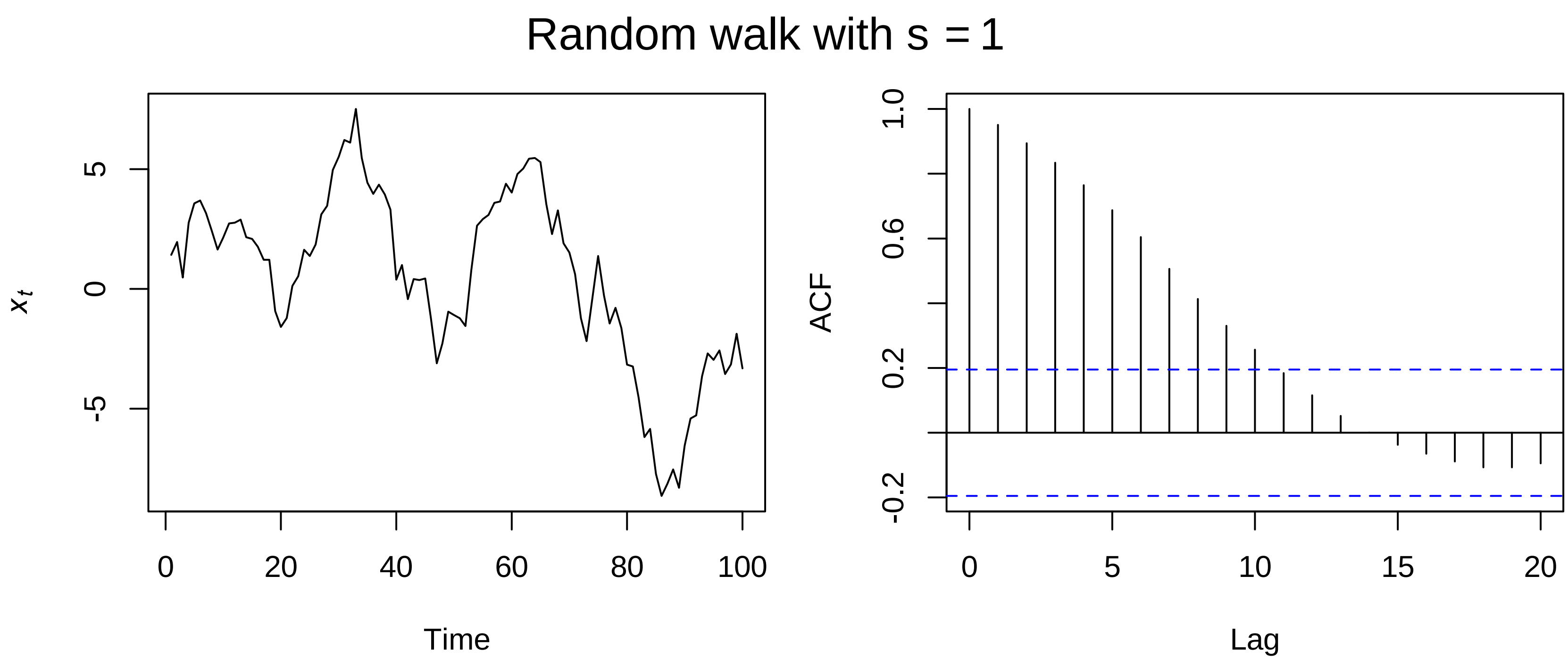
Classic time-series models

- A *time series model* for $\{x_t\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is thought to be a realization.

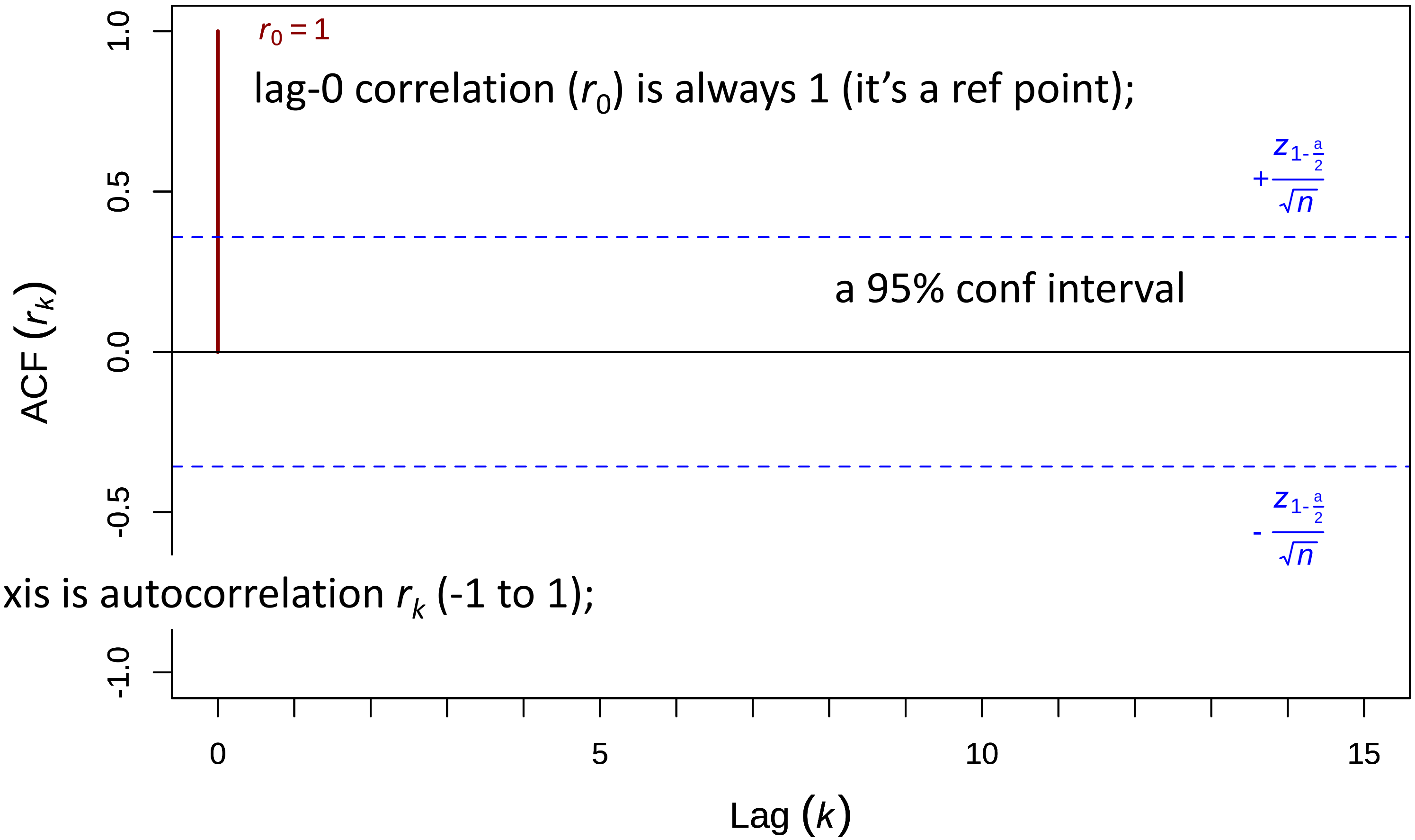
  - Time-series models – these describe the e_t
    - White noise
    - Autoregressive (AR) models
    - Moving average (MA) models
    - ARMA models
  - Random walks – an important type of non-stationary ts

Autocorrelation function (ACF): a powerful way to summarize a ts

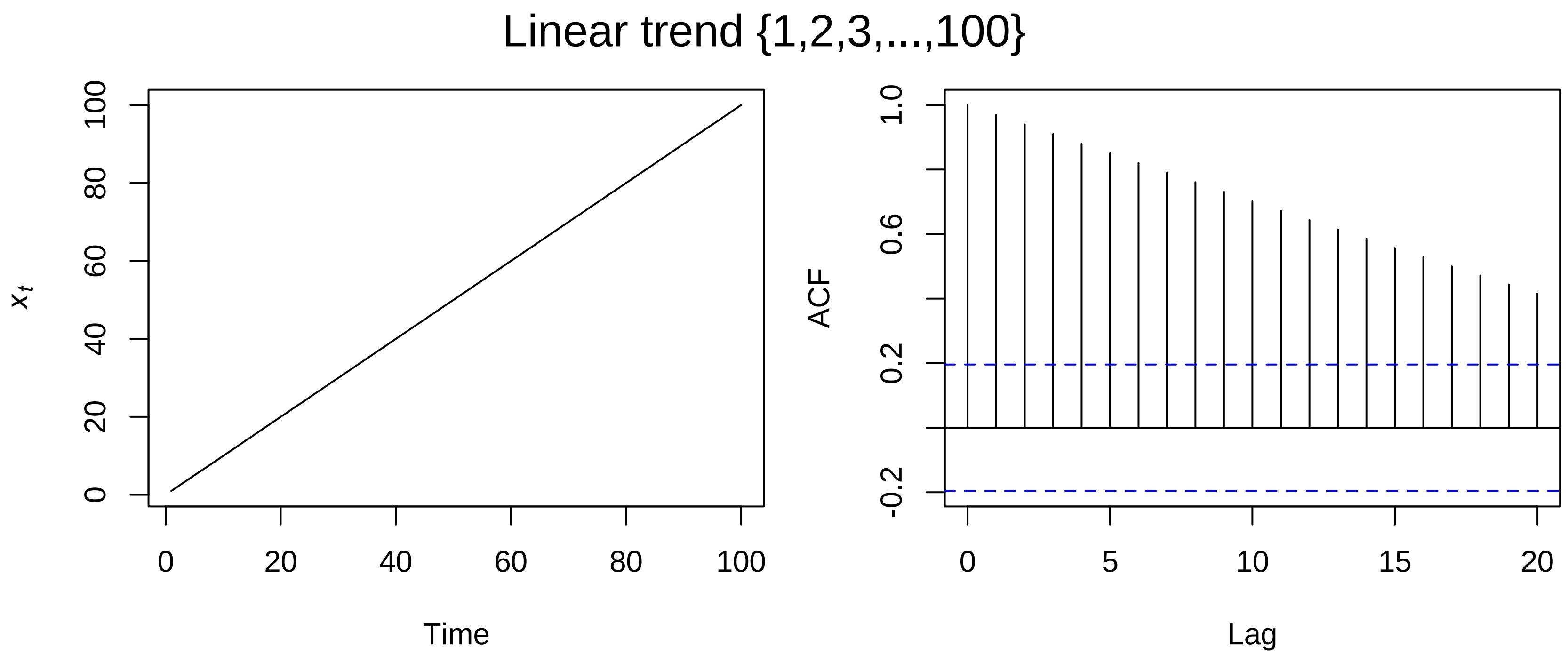
- ACF measures the correlation of a time series against a time-shifted version of itself (& hence “auto”)



The correlogram

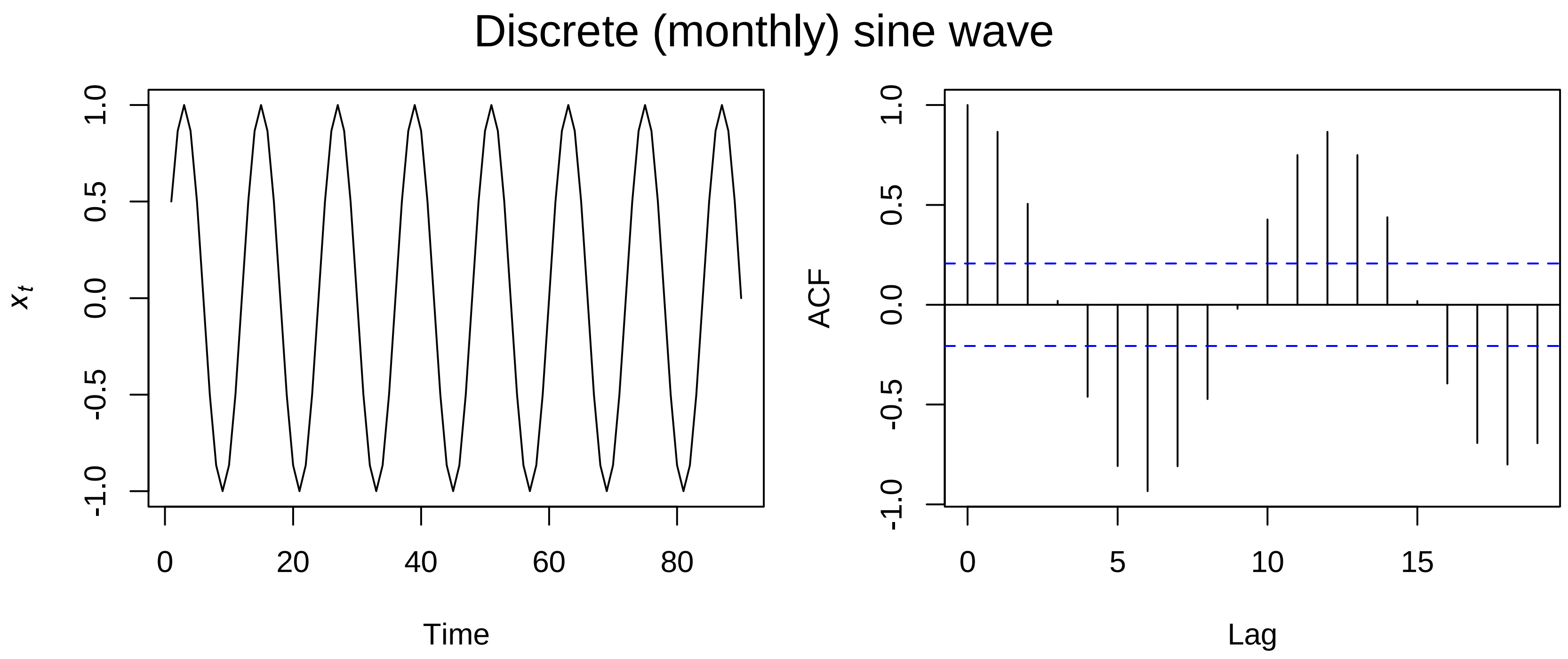


Correlogram for deterministic trend



Non-stationary

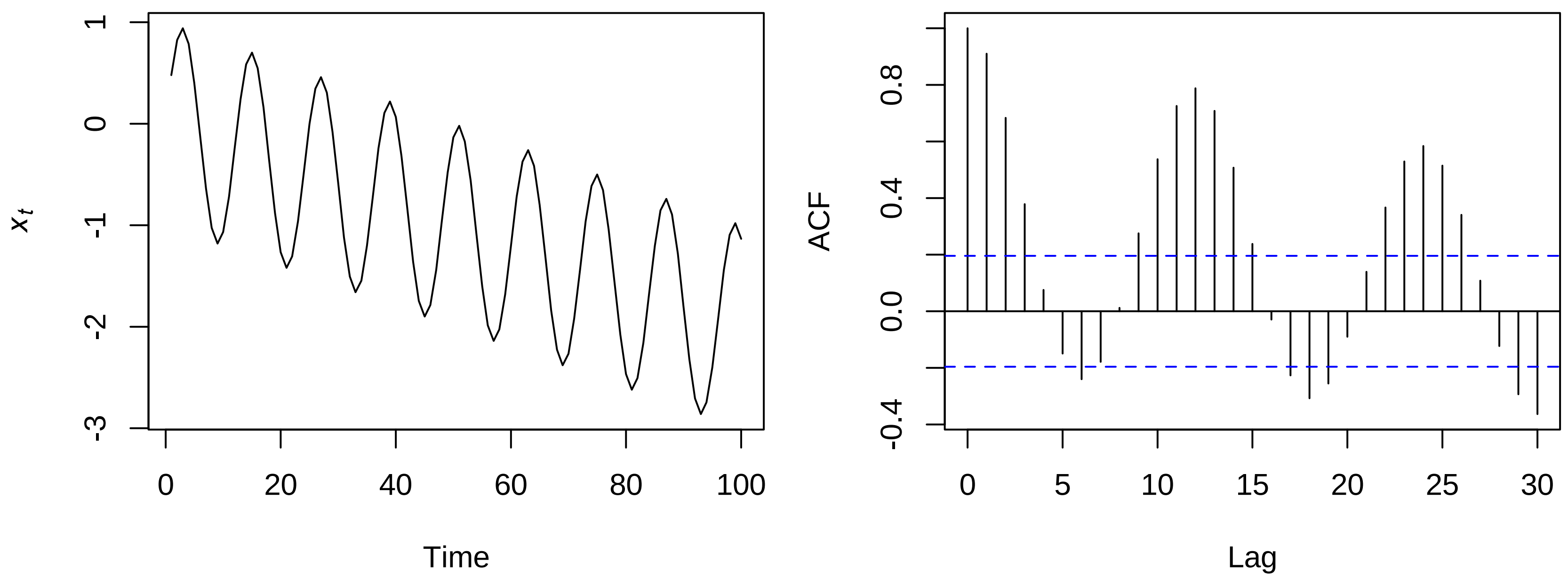
Correlogram for sine wave



Non-stationary

Correlogram for trend + season

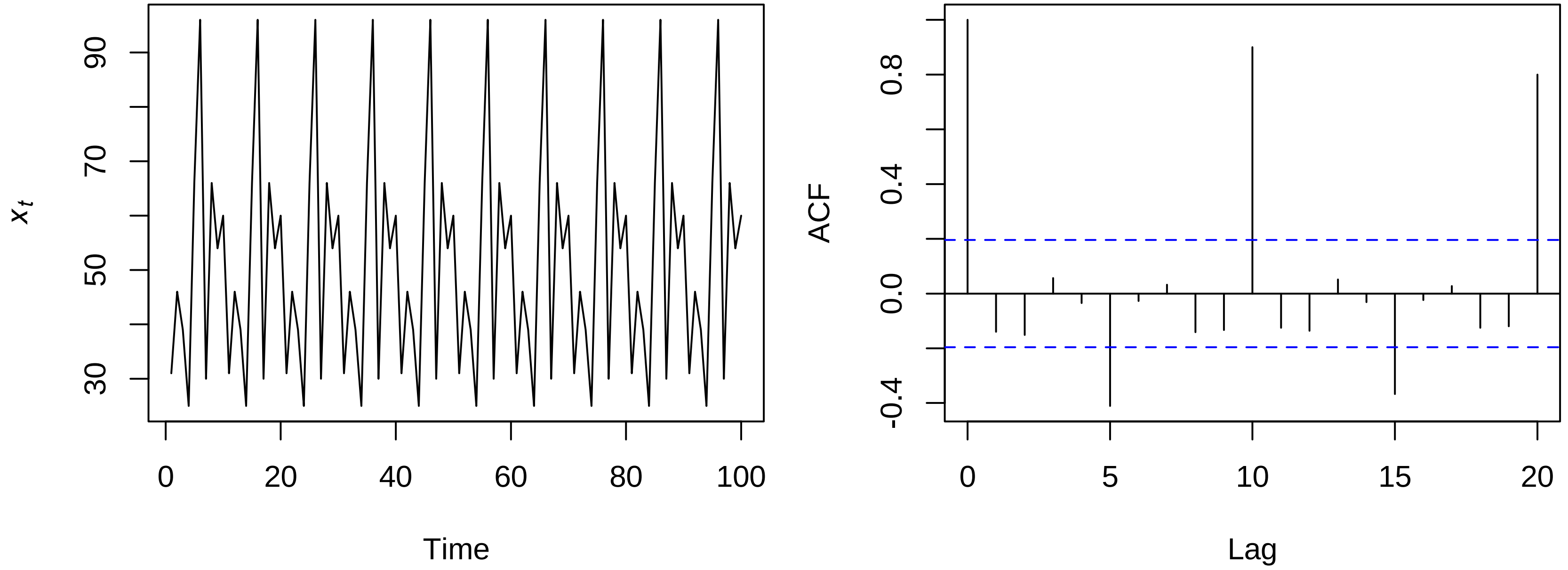
Linear trend + seasonal effect



Non-stationary

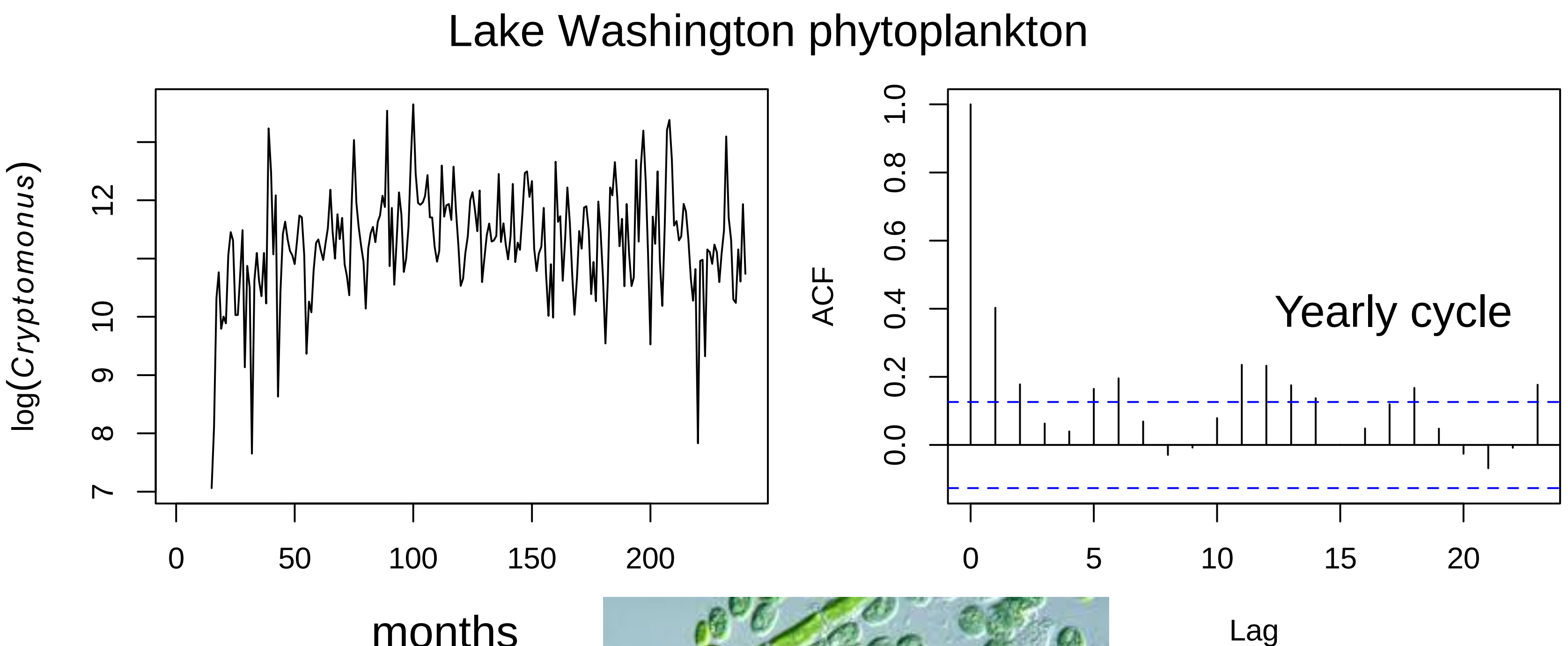
Correlogram for random sequence

Random sequence of 10 numbers repeated 10 times



Non-stationary

Correlogram for real data



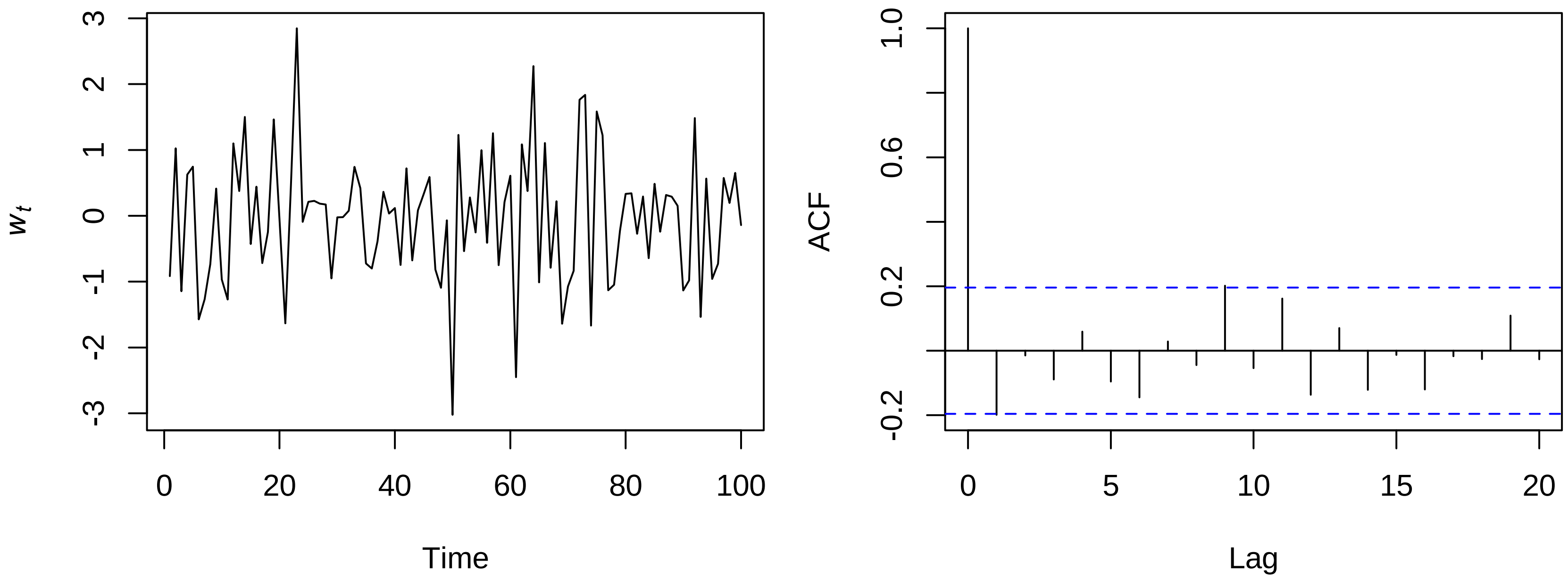
Classic time-series models

- A *time series model* for $\{x_t\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is thought to be a realization.

  - Time-series models – these describe the e_t
    - White noise
    - Autoregressive (AR) models
    - Moving average (MA) models
    - ARMA models
  - Random walks – an important type of non-stationary ts

White noise

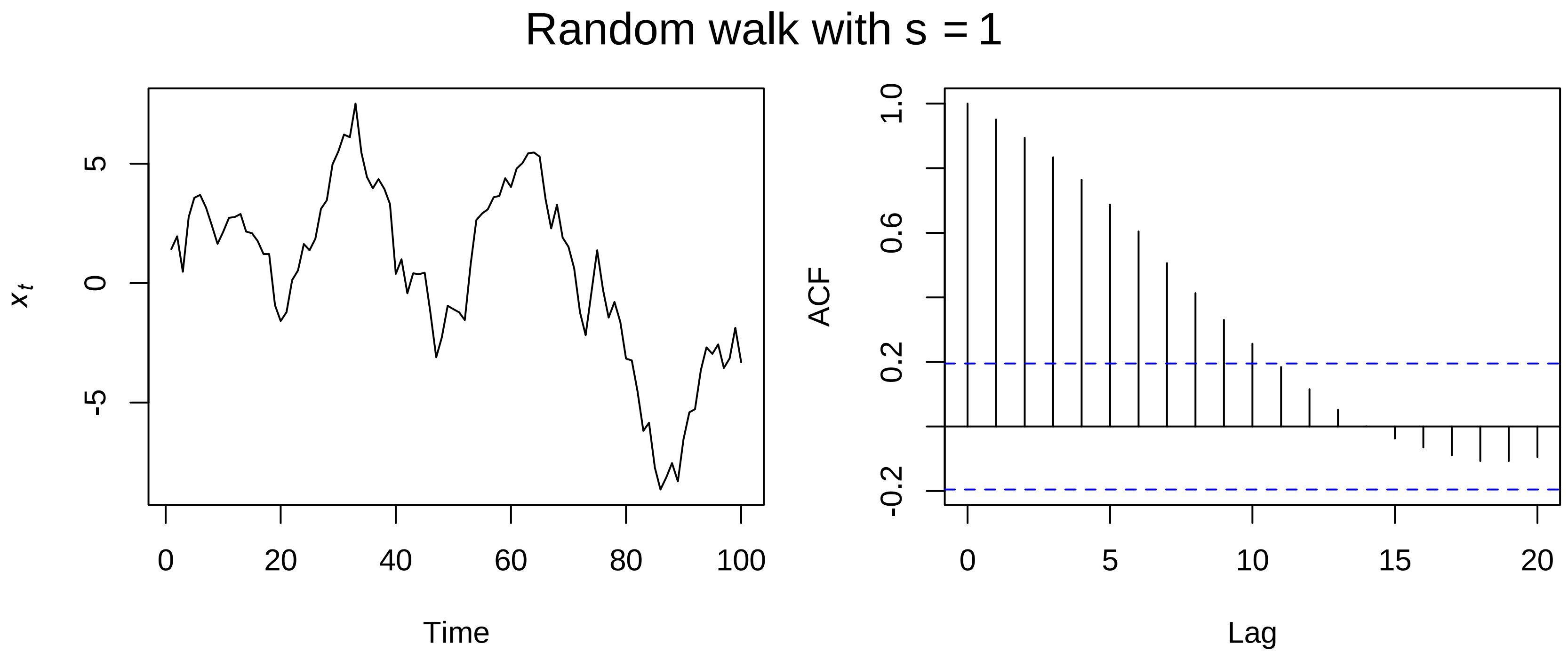
White noise with s = 1



white noise x_t are

1) *independent*, and
2) *identically distributed* with a mean of zero

Random walk (RW)



A time series $\{x_t : t = 1,2,3,\ldots,n\}$ is a *random walk* if

1) $x_t = x_{t\text{-}1} + w_t$, and
2) w_t is white noise

Random walks are NOT stationary!

Autoregressive (AR) models

- An *autoregressive* model of order *p*, or AR(*p*), is defined as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

  where we assume

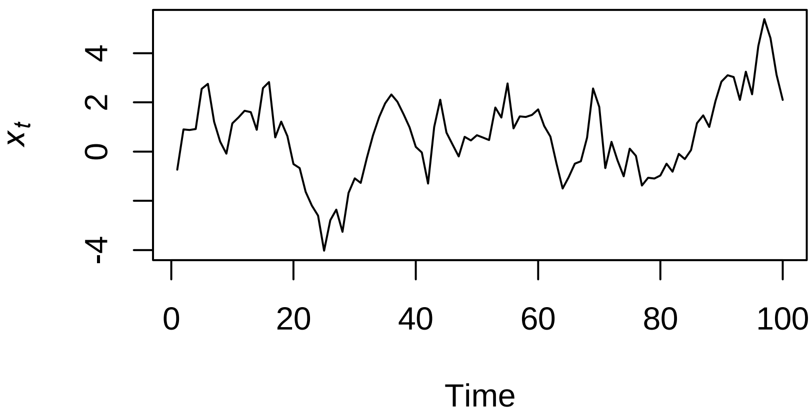
  1) w_t is WN, and

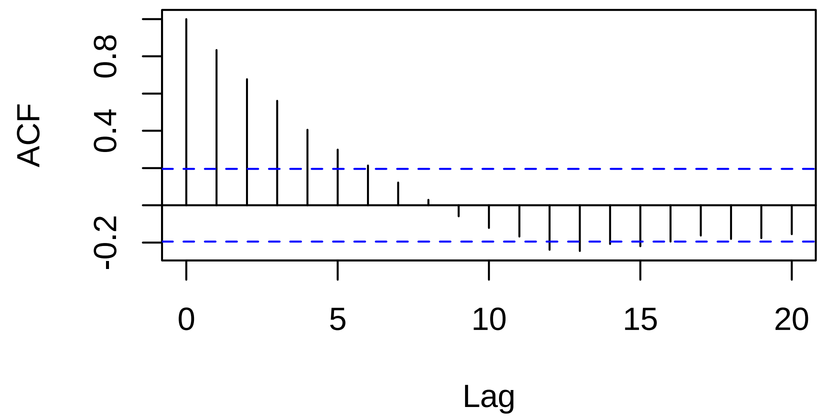
  2) $\phi_p \neq 0$ for order-*p* process

- *Note*: RW model is special case of AR(1) with $\phi_1 = 1$

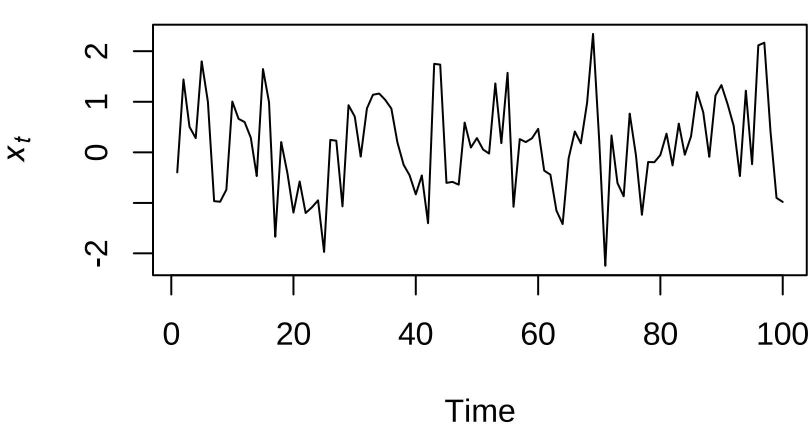
Examples of AR(1) processes

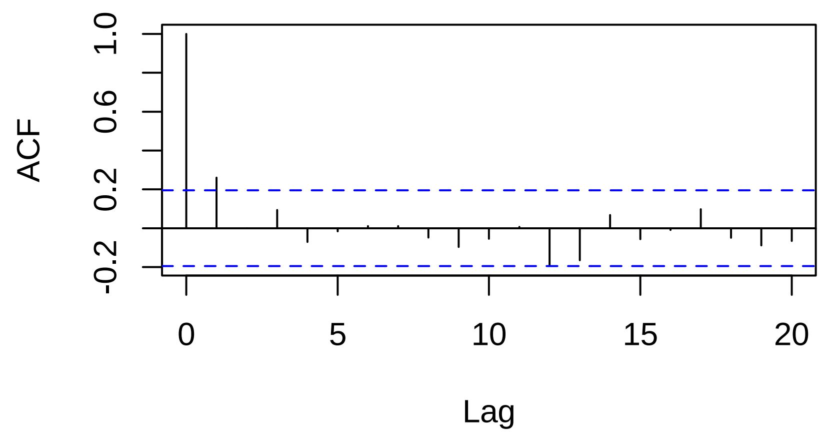
AR(1) with f = 0.9





AR(1) with f = 0.3

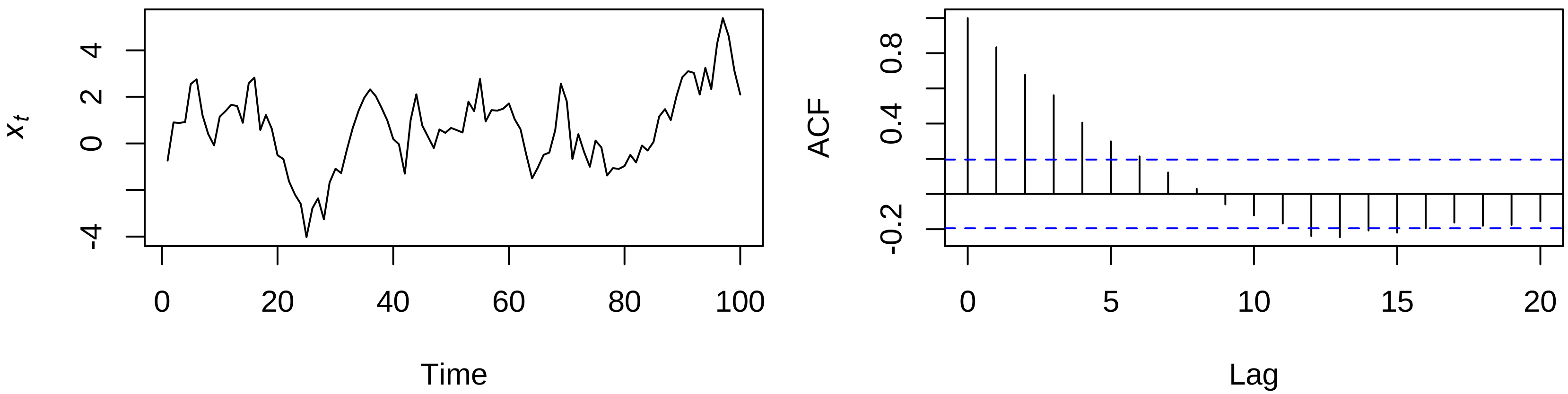




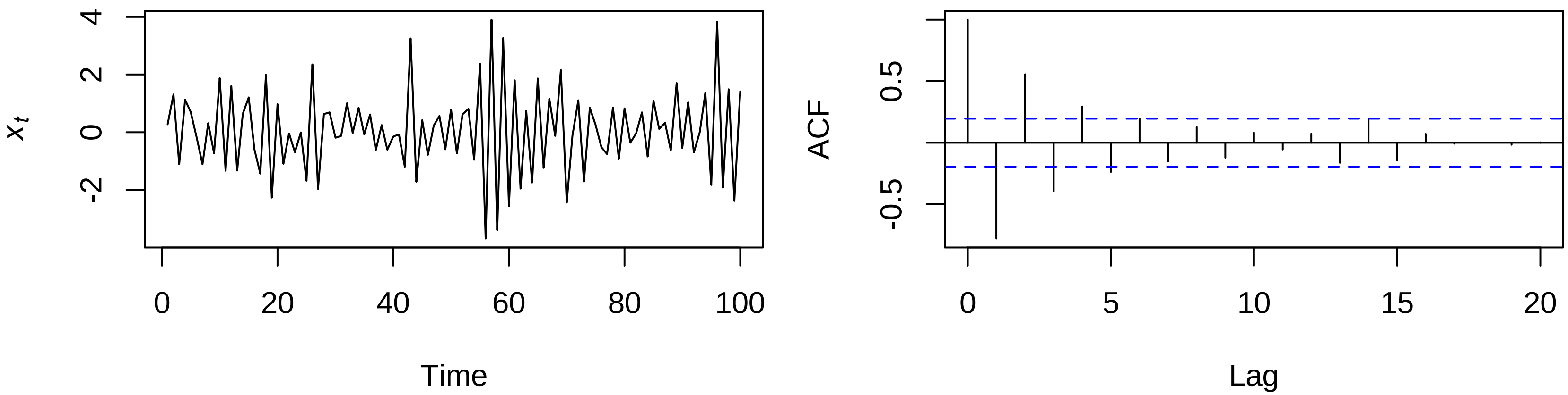
AR(p) can be stationary!

Examples of AR(1) processes

AR(1) with f = 0.9



AR(1) with f = - 0.9



Partial autocorrelation function

- The partial *autocorrelation function* (PACF) measures the linear correlation of a series x_t and x_{t+k} with the linear dependence of $\{x_{t-1},x_{t-2},\ldots,x_{t-(k-1)}\}$ removed

- It is defined as

$$\phi_{kk} = \begin{cases} \mathrm{Cor}(x_1, x_0) = \rho(1) & \text{if } k = 1 \\ \mathrm{Cor}\left(x_k - x_k^{k-1}, x_0 - x_0^{k-1}\right) & \text{if } k \geq 2 \end{cases} \qquad -1 \leq \phi_{kk} \leq 1$$

$$x_k^{k-1} = \beta_1 x_{k-1} + \beta_2 x_{k-2} + \cdots + \beta_{k-1} x_1$$

$$x_0^{k-1} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1}$$

Autoregressive (AR) models

- An *autoregressive* model of order p, or AR(p), is defined as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

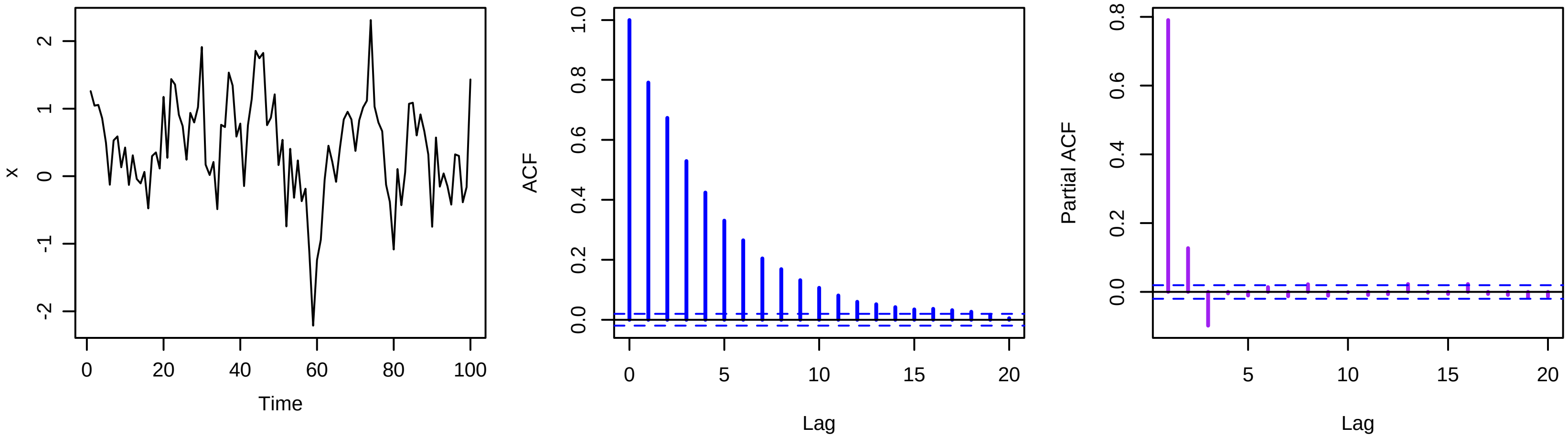
where we assume

1) w_t is WN, and
2) $\phi_p \neq 0$ for order-p process

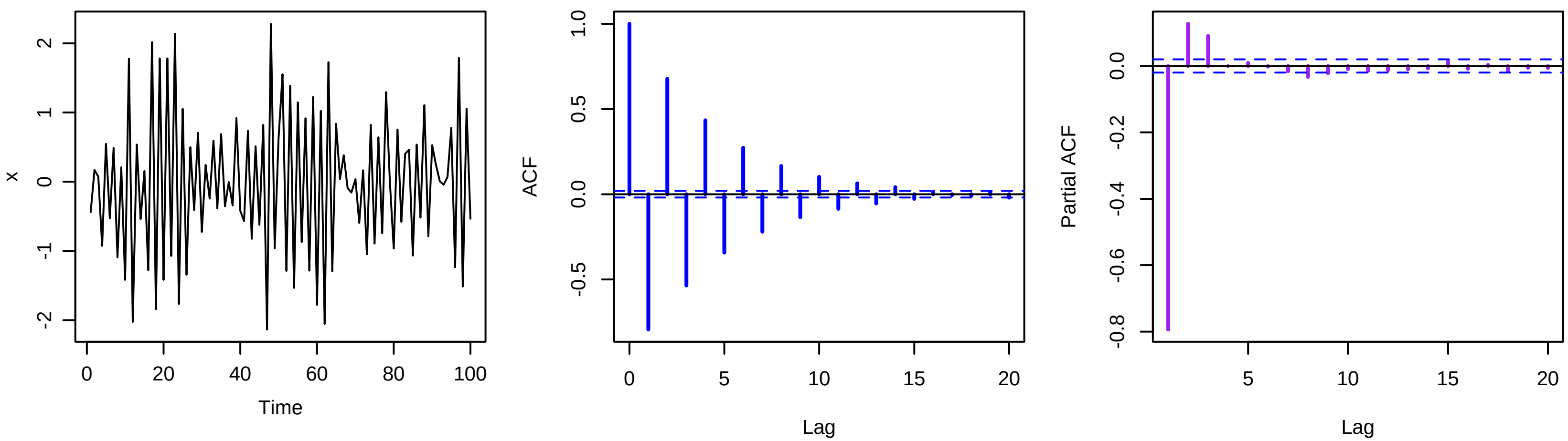
- *Note*: RW model is special case of AR(1) with $\phi_1 = 1$

ACF & PACF for AR(3) processes

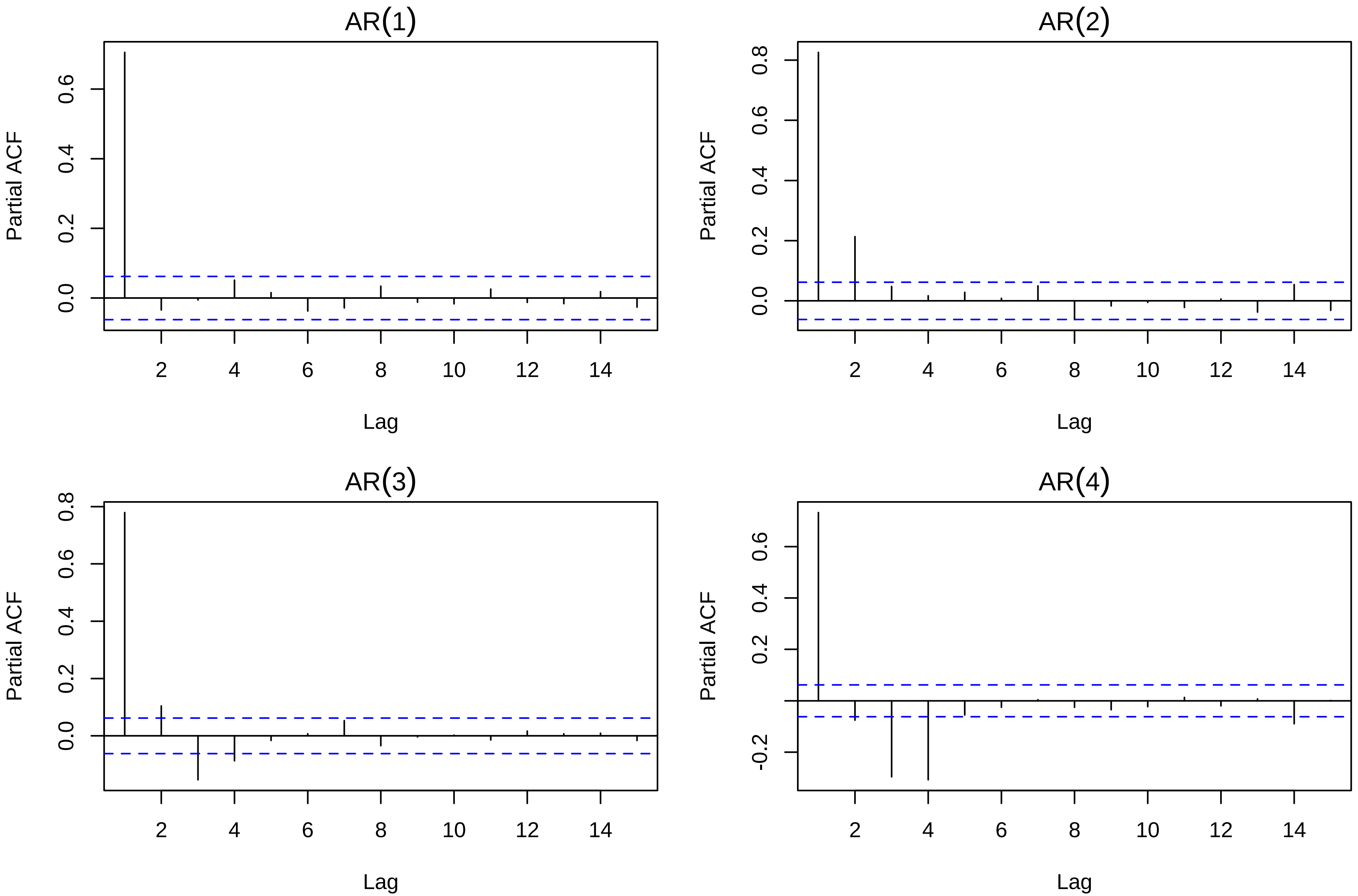
AR(3) with $f_1 = 0.7$, $f_2 = 0.2$, $f_3 = -0.1$



AR(3) with $f_1 = -0.7$, $f_2 = 0.2$, $f_3 = -0.1$



PACF for AR(p) processes



Moving average (MA) models

- A *moving average* model of order *q*, or MA(*q*), is defined as

$$x_t = w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

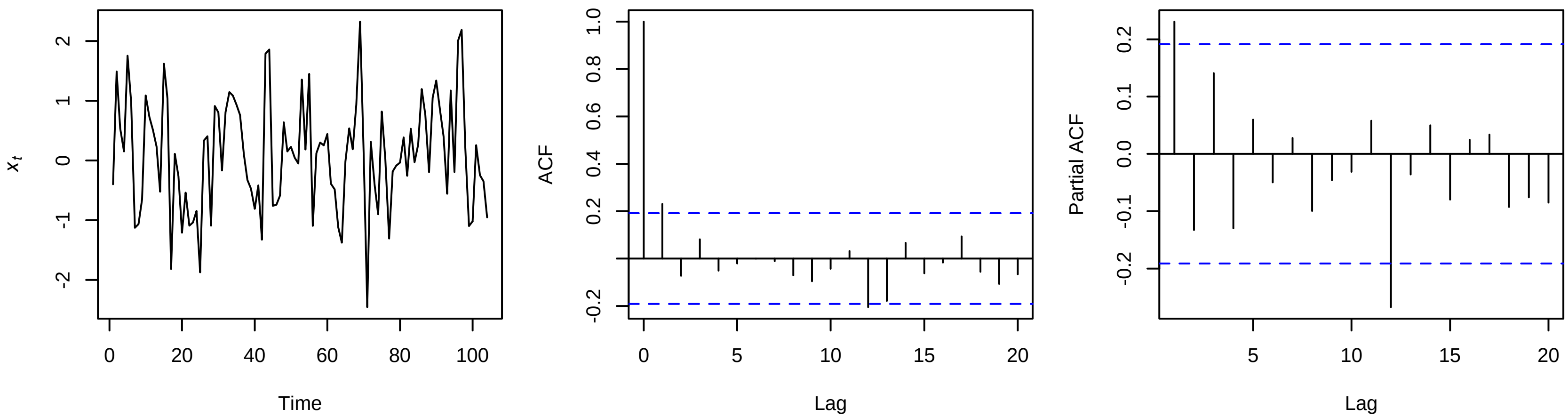
  where w_t is WN (with 0 mean)

- It is simply the current error term plus a weighted sum of the *q* most recent error terms

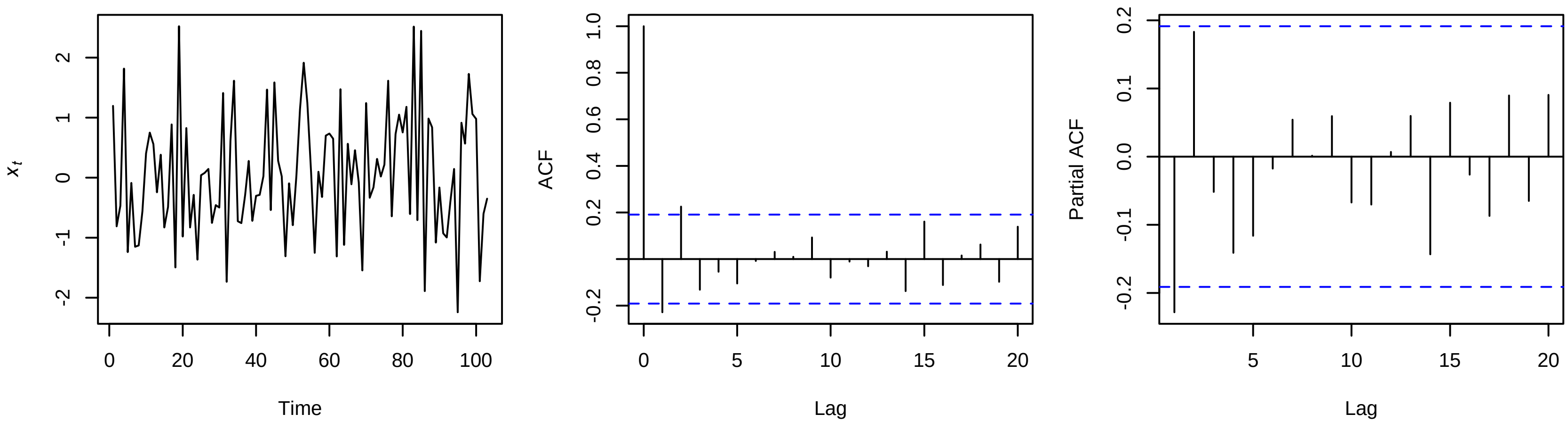
- Because MA processes are finite sums of stationary WN processes, they are themselves stationary

Examples of MA(q) processes

MA(1) with q = 0.3



MA(2) with q_1 = -0.3, q_2 = 0.3



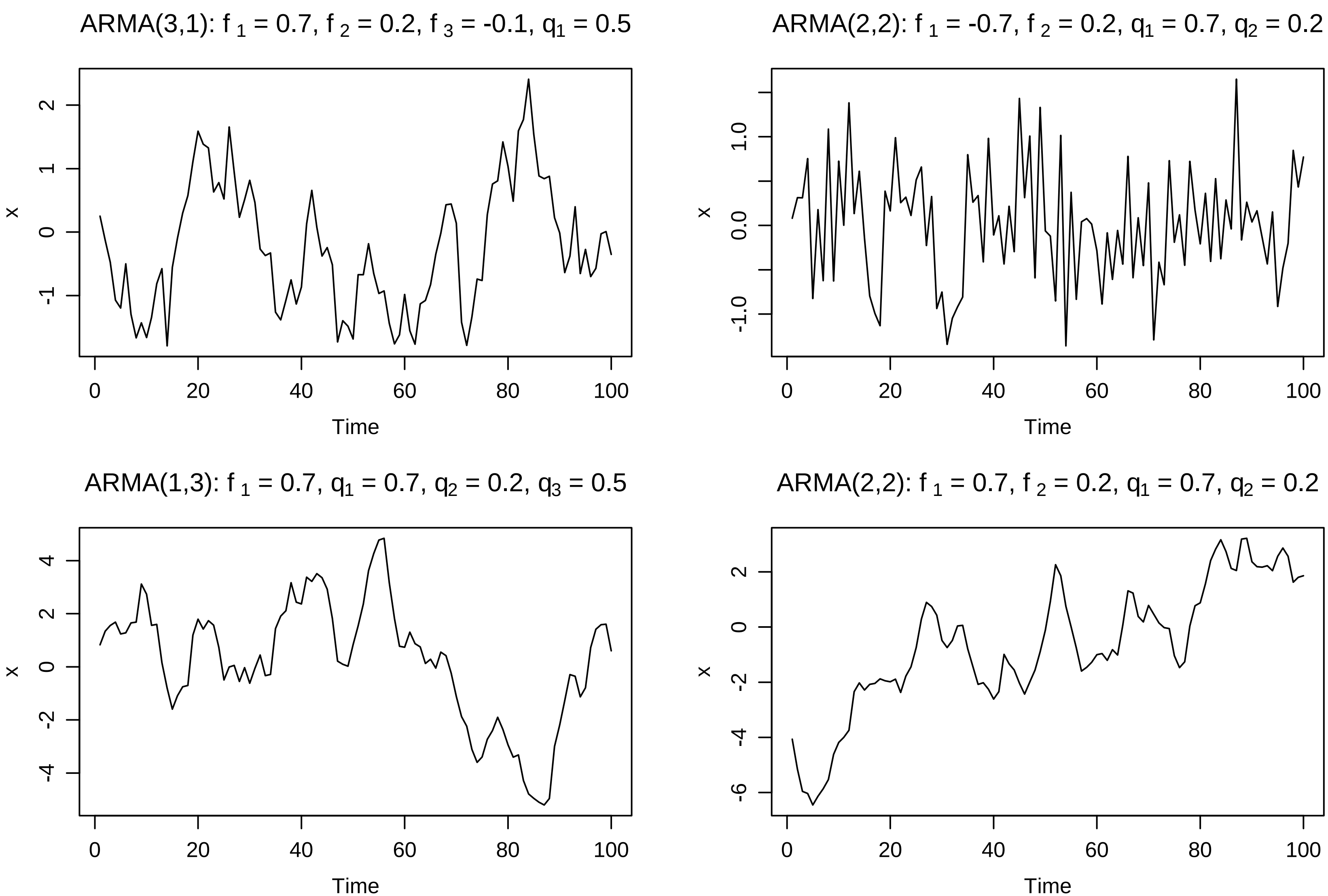
Autoregressive moving average models

- A time series is *autoregressive moving average*, or ARMA(p,q), if it is stationary and

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

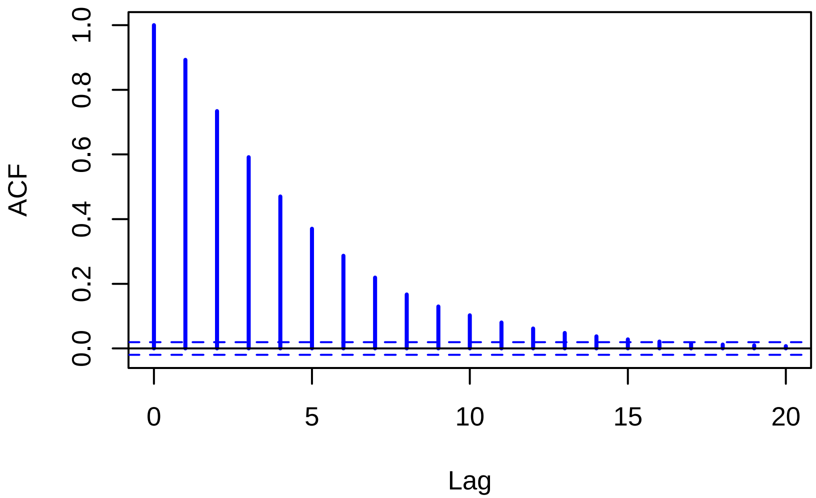
- Combines both AR(p) and MA(q)

Examples of ARMA(*p*,*q*) processes

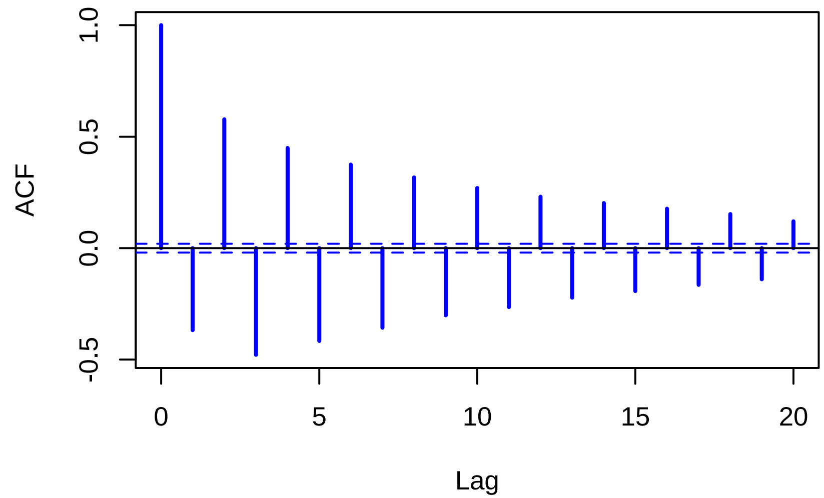


ACF for ARMA(*p*,*q*) processes

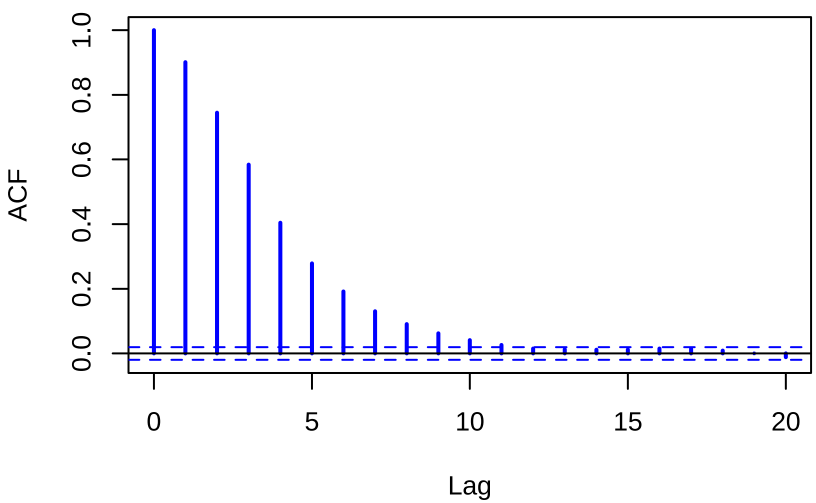
ARMA(3,1): $f_1 = 0.7$, $f_2 = 0.2$, $f_3 = -0.1$, $q_1 = 0.5$



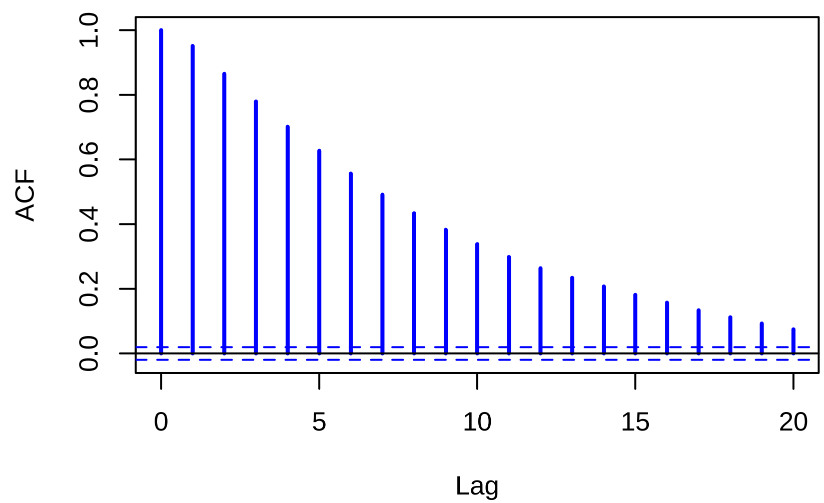
ARMA(2,2): $f_1 = -0.7$, $f_2 = 0.2$, $q_1 = 0.7$, $q_2 = 0.2$



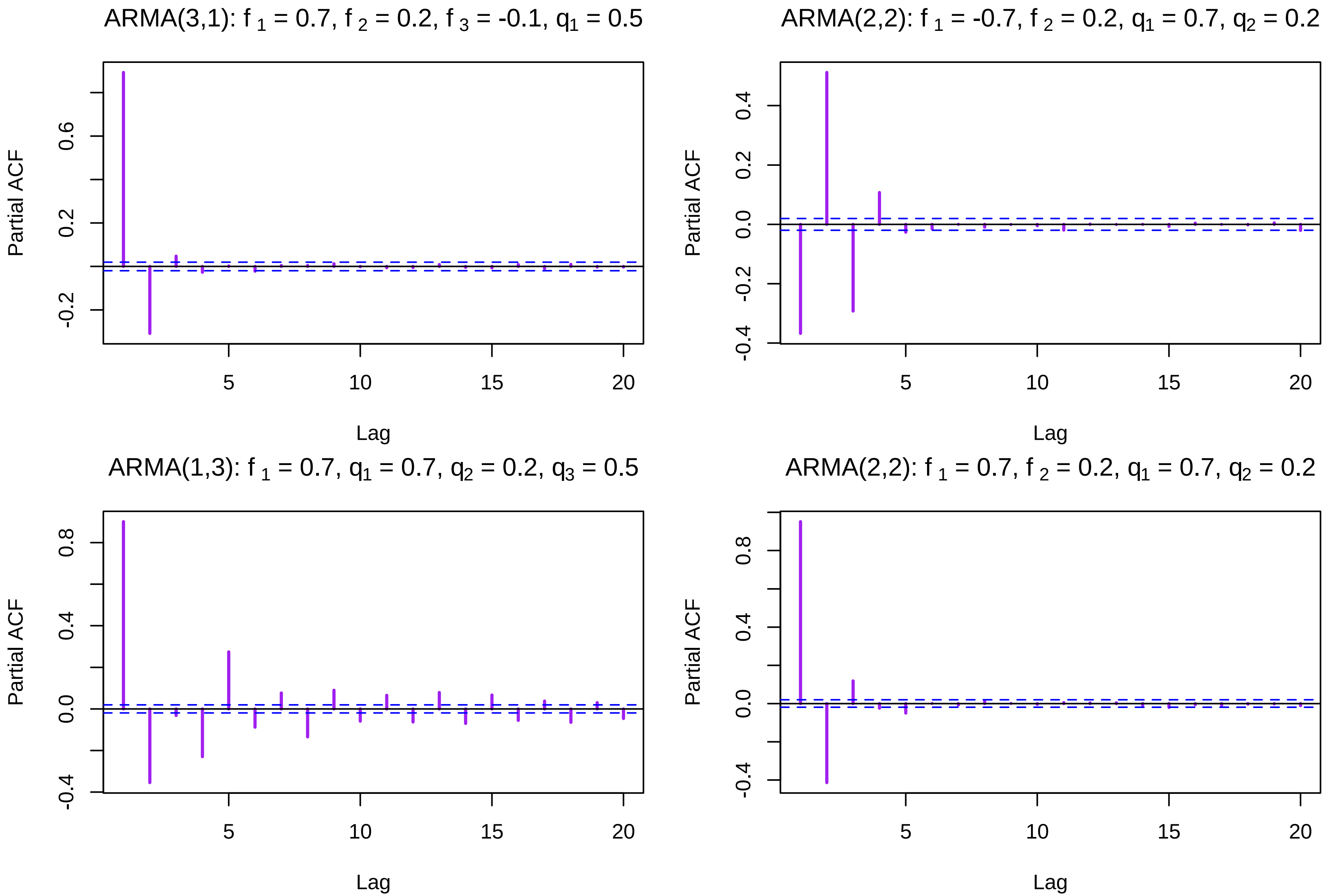
ARMA(1,3): $f_1 = 0.7$, $q_1 = 0.7$, $q_2 = 0.2$, $q_3 = 0.5$



ARMA(2,2): $f_1 = 0.7$, $f_2 = 0.2$, $q_1 = 0.7$, $q_2 = 0.2$



PACF for ARMA(*p*,*q*) processes



Difference to remove trend/season

- Differencing is a very simple means for removing a trend or seasonal effect
- The 1st-difference removes a linear trend, a 2nd-difference would remove a quadratic trend, etc.
- For seasonal data, using a 1st-difference with *lag = period* removes both trend & seasonal effects
- Pro: no parameters to estimate
- Con: no estimate of stationary process

Using ACF & PACF for model ID

	ACF	PACF
AR(p)	Tails off	Cuts off after lag-p
MA(q)	Cuts off after lag-q	Tails off
ARMA(p,q)	Tails off (after lag [q-p])	Tails off (after lag [p-q])

Topics for this lab

- `ts` class in R
- Plotting ts objects
- Understand covariance & correlation
- Examine some simple ts models
- Use `diff()` for trend/season removal
- Examine properties via `acf()` & `pacf()`
- Examine AR(p) models
- Examine MA(q) models
- ARMA(p,q) models via ‘arima.sim()’